Eric Slud October 30, 2020

STAT 770 Handout on Power for Tests in GLMs

This handout is about finding power and sample-size formulas, in Score and Waldtests about Generalized Linear Model parameters, based on local (contiguous) alternatives. The derivation of these formulas is based on somewhat advanced probability theory techniques, but the resulting formulas are directly applicable and interpretable by the applied statistician.

This topic is not covered in the Categorical Data Analysis book, except indirectly in talking about special power and sample-size formulas, for example in Sections 6.4 and 6.6. The technical derivations and formulas given in this handout can be used to make precise sense of assertions in Agresti, and we give a couple of examples of this, with another example to be worked out by students in an upcoming Homework Problem set.

Consider a Generalized Linear Model with fixed regressors $\{X_i\}_{i=1}^n \subset \mathbb{R}^p$, independent outcome data $Y_i \sim f(y, \theta_i) \equiv h(y) \exp(\theta_i y - c(\theta_i))$ leading to mean $\mu_i = c'(\theta_i)$ and variance function $v(\mu_i) = c''((c')^{-1}(\mu_i))$, and link function g, so that $g(\mu_i) = X_i'\beta$ for unknown parameter $\beta \in \mathbb{R}^p$. Here the parameters θ_i are assumed scalar, falling within the natural parameter interval which is assumed open (but possibly infinite). In this setting, the usual regularity conditions hold for large-sample asymptotic normality of MLE's and for asymptotic equivalence of Score and Wald tests. (Additional regularity conditions needed below for the local alternatives $H_{A,n}$ to have the *contiguity* propery are also satisfied in this case.) See the Handout **(3).** of the STAT 770 Course Web-page for details of the asymptotic relationships between MLEs and Wald and Score Test statisticss.

We consider a setting in which $\beta = (\gamma, \lambda)$, $\gamma \in \mathbb{R}^q$, $\lambda \in \mathbb{R}^{p-q}$, and it is desired to test $H_0 : \gamma = \mathbf{0}$ versus altenatives, and to derive sample-size formulas based on formulas for power of the Score Test of H_0 against nearby alternatives. The theory leading to such formulas is based on the notion of **contiguous alternatives**. This is a somewhat advanced topic that can be found in books like Bickel and Doksum's (1977) *Mathematical Stistics* book in simplified fom, and more fully in the advanced book *Asymptotic Statistics* (1998) by A. van der Vaart or the monograph *Semiparametric Theory and Missing Data* (2006) by A. Tsiatis.

For simplicity of assumptions and formulas, we restrict to the case where the underlying regressor variables X_i were themselves generated iid from a fixed (unchanging, known) probability law on $\mathbb{R}^p$, and where the link function g is *canonical*, which means that $\theta_i = X_i^{tr}\beta$, or equivalently means that $g^{-1}(y) \equiv c'(y)$. In that case, the family of alternatives $H_{A,n} : \beta = (0, \lambda) + b/\sqrt{n}$ for a fixed specified $b \in \mathbb{R}^p$ can be shown to be *contiguous* to $H_0 : \gamma = \mathbf{0}$, which means that any sequence Z_n of random variables that converge to 0 in probability as $n \to \infty$ under $P_{(\mathbf{0},\lambda_0)}$ (where $\beta_0 = (\mathbf{0}, \lambda_0)$ is a fixed parameter value within H_0) must also converge to 0 in probability under $P_{(0,\lambda_0)+b/\sqrt{n}}$. As a further item of notation, let b be decomposed into its initial q-dimensional and final $(p-q)$-dimensional subvectors as $b = (b_\gamma,\, b_\lambda)$.

A. Wald and Score Statistic Large-Sample Asymptotics

We start with some notation for Score Statistics and Observed and theoretical Information matrices. The Likelihood $L(\beta) \equiv L(\gamma, \lambda) = \prod_{i=1}^n f(Y_i, X_i^{tr}\beta)$ is the one given for data Y_i under the exponential-family model with density f, treating the X_i predictors as fixed. Let $\hat{\beta} = (\hat{\gamma}, \hat{\lambda})$ denote the unrestricted maximum likelihood estimator, which exists and is unique as a solution of $\nabla_\beta \log L(\beta) = 0$ because of the strict log-concavity of the exponential-family likelihood, with probability converging to 1 in large samples. Similarly, $\hat{\lambda}_r$ denotes the restricted maximum likelihood estimator of λ when γ is restricted to be equal to 0, i.e.,

$$\hat{\lambda}_r \qquad \text{solves} \qquad \nabla_\lambda \log L(0, \lambda) \; = \; \mathbf{0}$$

The Wald test statistic for $H_1 : \beta = \beta_0$ is given by $Z_n \equiv \hat{J}^{-1/2}\,(\hat{\beta} - \beta_0)$, and the Wald test statistic for $H_0 : \; \gamma = 0$ is given by

$$\tilde{Z}_n \; = \; \left\{(\hat{J}^{-1})_{\gamma\gamma}\right\}^{-1/2} \hat{\gamma} \; = \; (\hat{J}_{\gamma\gamma} - \hat{J}_{\gamma\lambda}\hat{J}_{\lambda\lambda}^{-1}\hat{J}_{\lambda\gamma})^{1/2}\,\hat{\gamma} \tag{1}$$

where $\hat{J} = J(\hat{\beta})$ is the usual *observed information matrix* and we define the general negative Hessian matrix of the log-likelihood by the block decomposition

$$J(\beta) \; = \; \begin{pmatrix} J_{\gamma\gamma} & J_{\gamma\lambda} \\ J_{\lambda\gamma} & J_{\lambda\lambda} \end{pmatrix} \; = \; -\nabla_\beta^{\otimes 2} \log L(\beta) \; = \; -\nabla_\beta^{\otimes 2} \log L(\gamma, \lambda)$$

where $J_{\gamma\gamma}$ is $p \times p$, $J_{\lambda\gamma} = J_{\gamma\lambda}^{tr}$ is a $(p-q) \times p$ matrix, and $J_{\lambda\lambda}$ is $(p-q) \times (p-q)$. The score test statistic for $H_0 : \gamma = \mathbf{0}$ is

$$S_n \; = \; (\tilde{J}_{\gamma\gamma} - \tilde{J}_{\gamma\lambda}\tilde{J}_{\lambda\lambda}^{-1}\tilde{J}_{\lambda\gamma})^{-1/2}\, \nabla_\gamma \log L(0, \hat{\lambda}_r) \tag{2}$$

where the restricted observed information matrix $\tilde{J} = J(0, \hat{\lambda}_r)$. With reference to the Asymptotics Handout `http://www.math.umd.edu/~slud/s701.S14/MLEpdim.pdf`, **(3)** on the course web-page, here are several useful statements about convergence in probability that are known from MLE theory to hold under the null hypothesis:

$$n^{1/2}\left(\hat{\beta} - \begin{pmatrix} 0 \\ \lambda_0 \end{pmatrix} - \hat{J}^{-1}\,\nabla_\beta \log L(0, \lambda_0)\right) \;\xrightarrow{P}\; \mathbf{0} \tag{3}$$

$$S_n \,-\, (\tilde{J}_{\gamma\gamma} - \tilde{J}_{\gamma\lambda}\tilde{J}_{\lambda\lambda}^{-1}\tilde{J}_{\lambda\gamma})^{-1/2}\left(\nabla_\gamma - \tilde{J}_{\gamma\lambda}\tilde{J}_{\lambda\lambda}^{-1}\nabla_\lambda\right)\log L(0, \lambda_0) \;\xrightarrow{P}\; \mathbf{0} \tag{4}$$

$$\sqrt{n}\,(\hat{\lambda}_r \,-\, \lambda_0) \,-\, \sqrt{n}\;\hat{J}_{\lambda\lambda}^{-1}\,\nabla_\lambda \log L(0, \lambda_0) \;\xrightarrow{P}\; \mathbf{0} \tag{5}$$

$$n^{-1}\,\tilde{J} - \mathcal{I}(0, \lambda_0) \;\xrightarrow{P}\; \mathbf{0} \;, \qquad n^{-1}\,\hat{J} - \mathcal{I}(0, \lambda_0) \;\xrightarrow{P}\; \mathbf{0} \tag{6}$$

Here J, $\hat{J}$, $\tilde{J}$ are forms of full-sample information and grow approximately proportionately to n, while $\mathcal{I}$ always denotes *per-observation Fisher Information* and (when X_i are *iid*) does not depend on n. In addition to these, or easily derived from (1), (4) and (6), is the asymptotic equivalence of the Wald and Score test statistics under H_0:

$$\tilde{Z}_n \,-\, S_n \;\xrightarrow{P}\; \mathbf{0} \tag{7}$$

Therefore, using the property mentioned above that $H_{A,n}$ is a family of *contiguous alternatives*, the same four statements (3)-(6) about convergence in probability hold also under the densities $f(y, X_i^{tr}\beta)$ when $\beta = (\mathbf{0}, \lambda_0) + b/\sqrt{n}$ under $H_{A,n}$. In the following manipulations with formulas we define, for ease of writing, the $q \times q$ matrix notation

$$D_\gamma = \mathcal{I}_{\gamma\gamma} - \mathcal{I}_{\gamma\lambda}\,\mathcal{I}_{\lambda\lambda}^{-1}\,\mathcal{I}_{\lambda\gamma} \,, \quad \mathcal{I}(0,\lambda_0) = \begin{pmatrix} \mathcal{I}_{\gamma\gamma} & \mathcal{I}_{\gamma\lambda} \\ \mathcal{I}_{\lambda\gamma} & \mathcal{I}_{\lambda\lambda} \end{pmatrix} = -\,E_{(\mathbf{0},\lambda_0)}\Big[\nabla_\beta^{\otimes 2} \log L(\mathbf{0},\lambda_0)\Big] \tag{8}$$

B. Simplified Statistic Expressions under Canonical GLM

Now we make these expressions more concrete and explicit using the formulas for log-likelihood and its gradient under the canonical-link GLM. The formula that we derived in class (see for example Lecture 12 Slide 5) for the gradient of GLM log-likelihood at a general parameter β is:

$$\nabla_\beta \log L(\beta) = \sum_{i=1}^n X_i \,\frac{Y_i - \mu_i}{g'(\mu_i)\,v(\mu_i)} \,, \qquad J(\beta) = -\sum_{i=1}^n X_i\,\nabla_\beta^{tr}\big(\frac{Y_i - \mu_i}{g'(\mu_i)\,v(\mu_i)}\big)$$

While we could have proceeded to develop asymptotic expressions for this general-link GLM setting, for simplicity we are restricting to the case of canonical link, with γ the initial q-dimensional subvector of the β parameter vector. Let $X_i = \binom{\xi_i}{\zeta_i}$ denote the decomposition of the p-vectors X_i into q-dimensional and $(p-q)$-dimensional sub-vectors, respectively ξ_i, ζ_i. Then, in our setting $g'(\mu)\,v(\mu) \equiv 1$, and

$$\nabla_\beta \log L(\beta) = \sum_{i=1}^n \begin{pmatrix}\xi_i \\ \zeta_i\end{pmatrix} (Y_i - \mu_{i,0}) \,, \qquad \tilde{J} = J(0,\hat\lambda_r) = \sum_{i=1}^n v(\hat\mu_{i,r})\,X_i\,X_i^{tr} \tag{9}$$

where

$$\mu_{i,0} = g^{-1}(X_i^{tr}\begin{pmatrix}0\\ \lambda_0\end{pmatrix}) = g^{-1}(\zeta_i^{tr}\,\lambda_0) \,, \qquad \hat\mu_{i,r} \equiv g^{-1}(X_i^{tr}\begin{pmatrix}0\\ \hat\lambda_r\end{pmatrix}) = g^{-1}(\zeta_i^{tr}\,\hat\lambda_r)$$

By (4), (5) and (7) together with these GLM expressions, under $H_{A,n}$ we obtain

$$\tilde{Z}_n = S_n + o_P(1) = \big[n\,D_\gamma\big]^{-1/2} \sum_{i=1}^n \big(\xi_i - \mathcal{I}_{\gamma\lambda}(\mathcal{I}_{\lambda\lambda})^{-1}\,\zeta_i\big)\,(Y_i - \mu_{i,0}) + o_P(1) \quad \text{as} \quad n \to \infty \tag{10}$$

and by (6), the blocks of $\mathcal{I}$ are estimating by n^{-1} times the corresponding blocks of $\hat{J}$ or $\tilde{J}$, where

$$\hat{J}_{\gamma\gamma} = \sum_{i=1}^n v(\hat\mu_{i,r})\,\xi_i\,\xi_i^{tr} \,, \quad \hat{J}_{\lambda\gamma} = \sum_{i=1}^n v(\hat\mu_{i,r})\,\zeta_i\,\xi_i^{tr} \,, \quad \hat{J}_{\lambda\lambda} = \sum_{i=1}^n v(\hat\mu_{i,r})\,\zeta_i\,\zeta_i^{tr}$$

C. Power and Sample-Size Formulas

We continue our reasoning by noticing that Y_i in the summation of (10) is centered at the value $\mu_{i,0}$ appropriate to H_0. To understand this expression asymptotically also under $H_{A,n}$, we consider these centered terms under $H_{A,n}$ with the aid of Taylor expansions, as follows:

$$E_{H_{A,n}}(Y_i) \;=\; g^{-1}(\zeta_i^{tr}\lambda_0 + b^{tr}\,X_i/\sqrt{n}) \;\approx\; \mu_{i,0} + (g^{-1})'(\mu_{i,0})\,b^{tr}\,X_i/\sqrt{n} \;+\; o_P(1/\sqrt{n})$$

So, since $(g^{-1})'(\mu) = (c')'(\mu) = c''(\mu) = v(\mu)$ in the canonical-link setting,

$$Y_i - \mu_{i,0} = Y_i - E_{H_{A,n}}(Y_i) \;+\; v(\mu_{i,0})\,b^{tr}\,X_i/\sqrt{n} \;+\; o_P(1/\sqrt{n}) \tag{11}$$

Combining (10) and (11), we find

$$S_n \;=\; D_\gamma^{-1/2}\,\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \mathcal{I}_{\gamma\lambda}\,(\mathcal{I}_{\lambda\lambda})^{-1}\,\zeta_i\right)\left\{\sqrt{n}(Y_i - E_{H_{A,n}}(Y_i)) \;+\; v(\mu_{i,0})\,X_i^{tr}\,b\right\} + o_P(1)$$

Since $Y_i - E_{H_{A,n}}(Y_i)$ have mean 0 and variance $c''(\zeta_i^{tr}\lambda_0 + b^{tr}X_i/\sqrt{n}) = v(\mu_{i,0}) + o_P(1)$ (uniformly in i, when with probability converging to 1 for large n, the vectors X_i are uniformly $o(\sqrt{n})$ for $1 \le i \le n$) under $H_{A,n}$, the definitions of $J, \mathcal{I}$ and (6) imply that under $H_{A,n}$

$$S_n \;=\; (D_\gamma)^{-1/2}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\xi_i - \mathcal{I}_{\gamma\lambda}(\mathcal{I}_{\lambda\lambda})^{-1}\,\zeta_i\right)(Y_i - E_{H_{A,n}}(Y_i)) \;+\; (D_\gamma)^{1/2}\,b_\gamma$$

where recall $b = \binom{b_\gamma}{b_\lambda}$, and thus

$$S_n \;\longrightarrow\; \mathcal{N}\Big(\,(D_\gamma)^{1/2}\,b_\gamma,\; I_{q\times q}\Big) \qquad \text{in distribution as} \qquad n \to \infty \tag{12}$$

The convergence in (12) quickly leads to power and sample-size formulas for hypothesis tests in problems related to our categorical data analysis topics. These tests are generally score tests or tests asymptotically equivalent to score tests, where in the test statistic is $T_n \equiv \|S_n\|_2 = \sum_{j=1}^q S_{n,j}^2$. The large-sample limiting power of such a test, which reduces to an ordinary two-tailed which rejects when $|S_n| \;\ge\; z_{\alpha/2}$, is known to be

$$P(\|S_n\|_2^2 \ge \chi^2_{q,\alpha}) \;=\; 1 - \texttt{pchisq}(\chi^2_{q,\alpha},\, q,\, b_\gamma^{tr}\,D_\gamma\,b_\gamma) \tag{13}$$

an upper-tail probability for the **noncentral chi-square** distribution with q degrees of freedom and *noncentrality parameter* $\|D_\gamma^{1/2}\,b_\gamma\|_2^2 = b_\gamma^{tr}D_\gamma\,b_\gamma$.

This formula allows one to determine the sample size needed to achieve a specified power against an alternative $\gamma_1 \in \mathbb{R}^q$. Suppose we are designing an experiment with known $\beta_1 = (\gamma_1, \lambda_1)$, D_γ, and want to know how large n must be to achieve power $\ge 1 - \delta$ against this fixed alternative. Let us assume that $\|\gamma_1\|_2 > 0$ is small, necessitating an n that is at least moderately largel. To find this n, we view $\lambda_0 = \lambda_1$ and $\gamma_1 = b_\gamma/\sqrt{n}$, $b = (b_\gamma,\, \mathbf{0})$, and equate

$$1 - \delta \;=\; 1 - \texttt{pchisq}\big(\chi^2_{q,\alpha},\, q,\; n\,\gamma_1^{tr}D_\gamma\,\gamma_1\big)\big) \tag{14}$$

The matrix D_γ would either be estimated by $n^{-1}\sum_{i=1}^{n} v(g^{-1}(\lambda_1^{tr}\zeta_i))\xi_i\,\xi_i^{tr}$ if the sequence $\{X_i\}_{i=1}^{n}$ is known, or else would be calculated as the expectation of this expression if the distribution of the *iid* variables X_i is known or assumed.

When n is large and $q = 1$, an approximate solution of (14) would be

$$1-\delta \;=\; 1-\Phi(z_{\alpha/2}-|\gamma_1\, D_\gamma^{1/2}|\,\sqrt{n}) + \Phi(-z_{\alpha/2}-|\gamma_1\, D_\gamma^{1/2}|\,\sqrt{n}) \;\approx\; 1-\Phi(z_{\alpha/2}-|\gamma_1\, D_\gamma^{1/2}|\,\sqrt{n})$$

and solve

$$\delta \;=\; \Phi(z_{\alpha/2}-|\gamma_1|\,\sqrt{n\,D_\gamma}) \quad\Longrightarrow\quad n \;\approx\; \big((z_\delta+z_{\alpha/2})\,/\,(D_\gamma^{1/2}\,\gamma_1)\big)^2 \tag{15}$$

We next consider the use of these power and sample-size formulas in a couple of examples.

D. Comparing Power of Trend versus General-Alternative Tests

D.1 *Score test in $K\times 2$ table*

Consider the data structure of Agresti's Sec. 5.3.7, in our notation for a $K\times 2$ table. Fix numbers of observations n_k, $k = 1,\ldots K$ and , and let $n = \sum_{k=1}^{K} n_k$. Fix a set of known nonnegative 'scores' x_k, increasing with respect to k, quantifying the (hypothesized) effect of a treatment, and $x_1 = 0$ without loss of generality. Let $A_k = \{n_1 + \cdots n_{k-1} + 1, \ldots,$ $n_1 + \cdots + n_k\}$ denote the index set of i's for which observations $Y_i \sim \mathrm{Binom}\big(1, \texttt{plogis}(a_0 + a_1 x_k)\big)$ are independent and $X_i = \binom{\xi_i}{1} = \binom{x_k}{1}$. This is a logistic regression model as in earlier sections, with $v(\mu_i) \equiv \mu_i(1-\mu_i)$, for which we consider a score test of the hypothesis $H_0 : a_1 = 0$. Here $\beta = (a_1, a_0)$, with $\gamma = a_1$, $\lambda = a_0$. Assuming that not all of the observations Y_i are the same (i.e., not all 0 and not all 1), it is easy to check that the restricted MLE for a_0 puts

$$\hat{a}_{0,r} \;=\; \texttt{logit}(\bar{Y}) \;, \qquad \hat{\mu}_{i,0} \;\equiv\; \bar{Y} \;, \qquad \bar{Y} \;=\; n^{-1}\textstyle\sum_{i=1}^{n} Y_i$$

From this, it follows by definition that

$$\tilde{J} \;=\; \bar{Y}(1-\bar{Y})\sum_{i=1} X_i X_i^{tr} \;=\; \bar{Y}(1-\bar{Y})\sum_{k=1}^{K} n_k \begin{pmatrix} x_k \\ 1 \end{pmatrix}\begin{pmatrix} x_k \\ 1 \end{pmatrix}^{tr}, \quad D_\gamma \;=\; \frac{\bar{Y}(1-\bar{Y})}{n}\sum_{k=1}^{K} n_k\,(x_k - \bar{\xi})^2$$

where $\bar{\xi} \;\equiv\; n^{-1}\sum_{i=1}^{n}\xi_i \;=\; n^{-1}\sum_{k=1}^{K} n_k\,x_k$. Therefore the score statistic S_n in (2) is directly calculated equal to the Cochran-Armitage Trend Test statistic $\sum_{i=1}^{n}(\xi_i - \bar{\xi})\,\cdot$ $(Y_i - \bar{Y})\,/\,\big[\sum_{k=1}^{K} n_k\,(x_k - \bar{\xi})^2\big]^{1/2}$ which is given as (5.7) and discussed in the present context at the top of p. 179.

We find the power of this Score Test against two specific contiguous alternatives.

(i) First: we consider the alternatives $H_{A,n} : a_1 = a/\sqrt{n}$. The power of the score test against $H_{A,n}$ is given by (12) or (13) as

$$1 - \texttt{pchisq}(\chi^2_{1,\alpha},\, 1,\, a^2\,D_\gamma) \;=\; 1 - \Phi(z_{\alpha/2} - |a|\,D_\gamma) + \Phi(-z_{\alpha/2} - |a|\,D_\gamma) \tag{16}$$

(ii) Second, we consider local alternatives to H_0 that show single-k differences from nullity rather than trend, such as $H_{B,n}: \ \pi_k = \texttt{plogis}\big(a_0 + h\, I_{[k=1]}/\sqrt{n}\big)$. The score-statistic power formula (13) does not apply directly, but the expression of S_n in the Cochran-Armitage form does. Recall that A_k is the index-set of i's for which $n_1 + \cdots + n_{k-1} < i \le n_1 + \cdots + n_k$, so that $A_1 = \{1, \ldots, n_1\}$. Using (11), we find that

$$E_{H_{B,n}}(Y_i - \bar{Y}) \;=\; (h/\sqrt{n})\,\big(I_{[i \le n_1]} - \frac{n_1}{n}\big) \;+\; o(1/\sqrt{n})$$

so that the conditional expectation of S_n given $\{X_i\}_{i=1}^n$ is

$$E_{H_{B_n}}\big(S_n \,|\, \{X_i\}_{i=1}^n\big) \;=\; \big(\frac{h}{\sqrt{n}}\,(n_1(x_1 - \bar{\xi})\big)\;/\;\Big[\sum_{k=1}^K n_k\,(x_k - \bar{\xi})^2\Big]^{1/2} \;+\; o(1/\sqrt{n}) < 0$$

Here we have used $x_1 = 0$, and since the variance of the Cochran-Armitage statistic under $H_{B,n}$ is still approximately 1 as $n \to \infty$, we conclude that the power of S_n against alternatives in $H_{B,n}$ is actually less than α.

D.2 *General-alternative X^2 test in $K \times 2$ table*

If, in the setting of Section D.1, $y_{k1} \equiv \sum_{i \in A_k} Y_i$ and $y_{k0} \equiv n_k - y_{k1}$, then the data y_{kj} are in the form of a $K \times 2$ table with two ordinal factors, but Y_i still obeys a canonical-link GLM. We compute the local (contiguous-alternative) power against alternatives of a size-α score test of $H_0: \ \pi_1 = \pi_2 = \cdots = \pi_K$. This test, with the same null-hypothesis but a richer set of alternatives, may be viewed as the score test for $\beta = (\gamma, a_0)$, within the saturated model where $\gamma = (\gamma_1, \ldots, \gamma_{K-1})$ satisfy $\pi_i = \texttt{plogis}\big(\gamma^{tr}(I_{[k=2]}, \ldots, I_{[k=K]}) + a_0\big)$ for $i \in A_k$. The associated regression variables are $X_i = (\xi_i, \zeta_i) = (I_{[k=2]}, \ldots, I_{[k=K]}, 1)$ (with $\zeta_i \equiv 1$) for $i \in A_k$. Here we find under $H_0: \ \gamma = \mathbf{0}$ that $X_i^{tr}\beta = a_0$ and $\mu_i = \mu_0 = \texttt{plogis}(a_0)$ for all i, and

$$J(\beta) \;=\; \mu_0(1-\mu_0)\begin{pmatrix} n_2 & \mathbf{0}^{tr} & 0 & n_2 \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ 0 & \mathbf{0}^{tr} & n_K & n_K \\ n_2 & \cdots & n_K & n \end{pmatrix},$$

$$D_\gamma \;=\; \mu_0(1-\mu_0)\left\{\begin{pmatrix} n_2 & \mathbf{0}^{tr} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0}^{tr} & n_K \end{pmatrix} - \frac{1}{n}\begin{pmatrix} n_2 \\ \vdots \\ n_K \end{pmatrix}\begin{pmatrix} n_2 \\ \vdots \\ n_K \end{pmatrix}^{tr}\right\}$$

In addition, again assuming that $\bar{Y} \ne 0, 1$, the restricted MLE of a_0 is again $\hat{a}_{0,r} = \texttt{logit}(\bar{Y})$, so that $\hat{\mu}_{i,r} \equiv \bar{Y}$.

We now consider the behavior of Score statistic (2) of H_0 versus contiguous alternatives of the form $H_{C,n}: \ \beta = (b_\gamma/\sqrt{n},\, a_0)$. As long as all ratios n_k/n tend to positive limits as $n \to \infty$, the score statistic is asymptotically equivalent to the display preceding (12), i.e. to

$$D_\gamma^{-1/2}\,\frac{1}{\sqrt{n}}\sum_{k=1}^K \begin{pmatrix} I_{[k=2]} - n_2/n \\ \vdots \\ I_{[k=K]} - n_K/n \end{pmatrix}(y_{k1} - n_k\bar{Y})$$

Then the score test $T_n = \|S_n\|_2^2$ is asymptotically χ^2_{K-1} distributed under H_0, and can be seen using (12) to be noncentral chi-square distributed under $H_{C,n}$ with degrees of freedom $K-1$ and noncentrality parameter

$$n^{-1}\, b_\gamma^{tr}\, D_\gamma\, b_\gamma \;=\; n^{-1}\, \mu_0(1-\mu_0)\, \Big[\sum_{k=2}^{K} n_k\, b^2_{\gamma,k-1} \;-\; n^{-1} \Big(\sum_{k=2}^{K} n_k\, b_{\gamma,k-1} \Big)^2 \Big]$$

We can apply this limiting distributional result to find the power of the score test in this section versus the alternatives (i) and (ii) considered in subsection D.1. In (i), $b_{\gamma,k-1} = a x_k$, for $k = 2, \ldots, K$, while in (ii), the alternative is essentially $b_\gamma = -(1,1,\ldots,1)\, h$. Thus the power for the test based on T_n versus alternatives (i) is

$$1 \;-\; \texttt{pchisq}\Big(\chi^2_{K-1,\alpha},\; K-1,\; a^2 \mu_0(1-\mu_0) \frac{1}{n} \sum_{k=1}^{n} n_k (x_k - \bar{X})^2\Big)$$

and the power versus alternatives (2) is

$$1 \;-\; \texttt{pchisq}\Big(\chi^2_{K-1,\alpha},\; K-1,\; h^2 \mu_0(1-\mu_0) \frac{n_1(n-n_1)}{n^2}\Big)$$